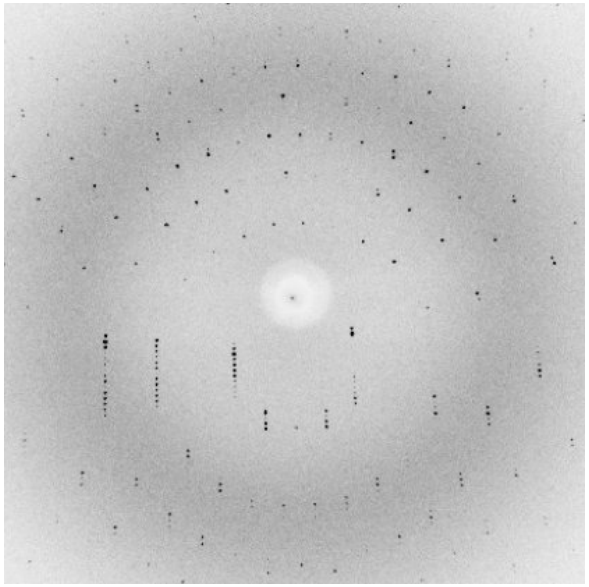
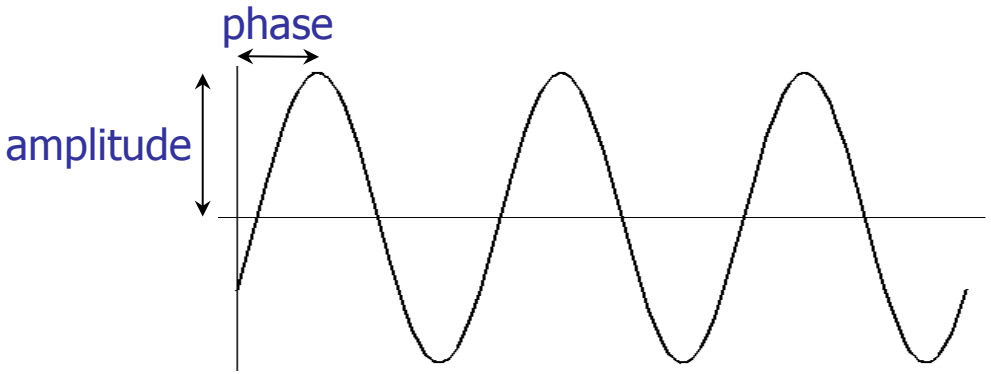
Molecular replacement

New structures from old

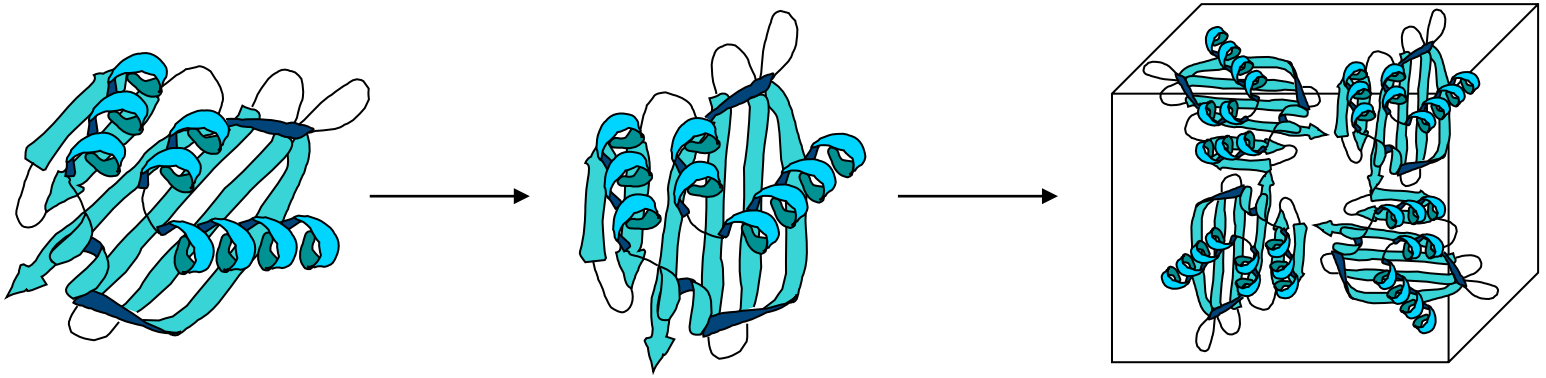
The Phase Problem





Phasing by molecular replacement

- Phases can be calculated from atomic model
- Rotate and translate related structure



Models for molecular replacement

- Ease of molecular replacement depends on quality and completeness of model
- Quality depends on:
  - sequence identity
  - experiment (resolution, NMR/X-ray)
  - flexibility
  - sophistication of homology modelling
- Maximum likelihood methods can cope with poorer models

Choosing the right model

- The best model may not be the top hit
  - correlation between sequence identity and quality is approximate
  - conformational change
- Test multiple choices of model
  - easier in a pipeline: phenix.MRage, Balbes, MrBUMP
  - multiple templates,
    multiple manipulations of template

Model manipulation

- Ensembler
  - multiple structure superposition to make ensemble of possible models
  - optionally trim non-conserved surface loops
- Sculptor (see also Chainsaw)
  - use sequence alignment to:
    - trim parts of template not in target
    - adjust B-factors of poorly-conserved regions
  - use surface accessibility to:
    - adjust B-factors of surface regions

Homology modeling and MR

- *Rosetta*: sophisticated modeling program from David Baker's group
  - computationally intensive (Rosetta@home)
  - combination of physics, database knowledge and conformational search algorithms
- Templates from NMR structures and distant homologues can be improved for MR
  - Bin Qian, Rhiju Das *et al.* (2007)
- Complete (possibly ambiguous) solution from poor model: phenix.mr_rosetta
  - Frank diMaio, Tom Terwilliger *et al.* (2011)

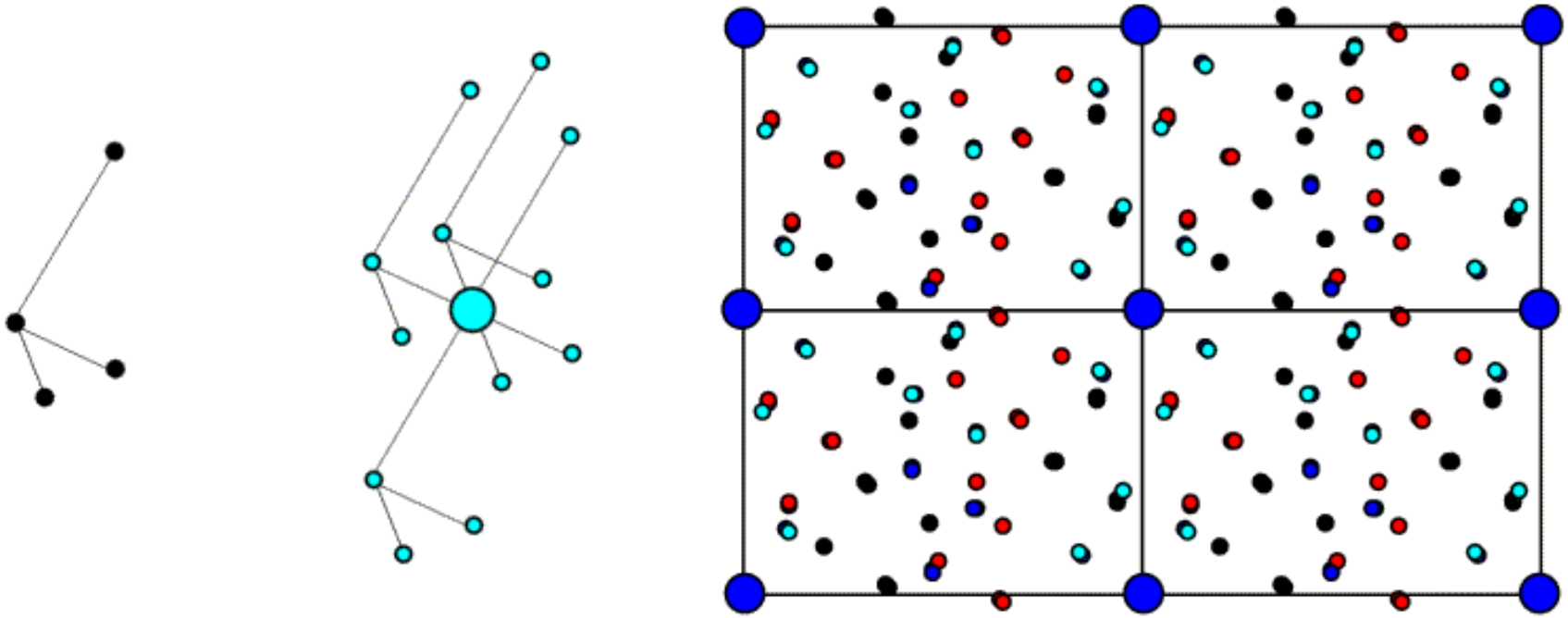
Ab initio modeling and MR

- CASP7: Baker group generated exceptionally good model for T0283
- 3 of top 5 models give clear MR solution
- complete automatically with rebuilding software

- AMPLE: CCP4 program

Patterson-based molecular replacement

- Original MR algorithms were based on properties of the Patterson map
  - discussed in detail on CIMR course web page

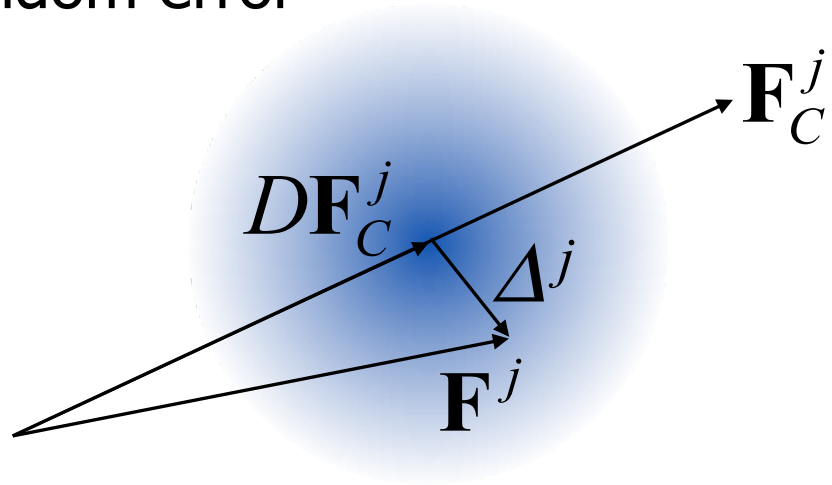


Likelihood-based molecular replacement

- Likelihood target:
  - probability of observed amplitude given (set of) model structure factor contributions
    - account for effect of unknown relative phases
- Benefits of likelihood
  - account for expected size of errors in model
  - account for lack of completeness of model
  - exploit knowledge from partial solutions
  - allow ensemble of possible models
    - also useful for MR with NMR

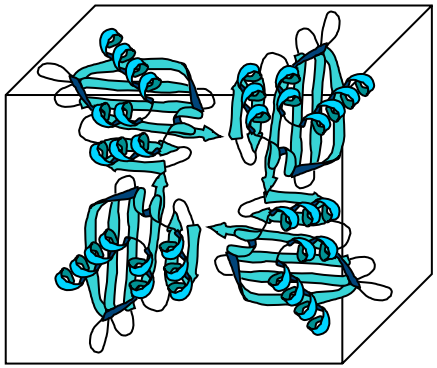
Effect of errors on structure factor distribution

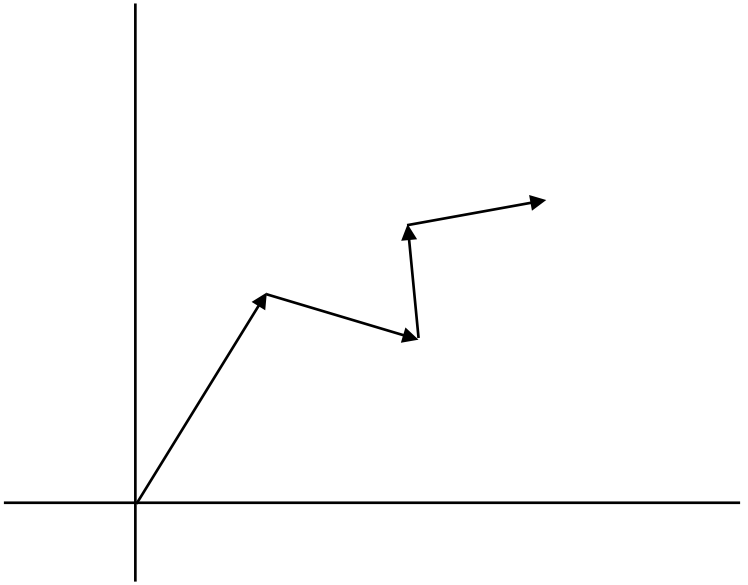
- Errors and incompleteness in model lead to errors in calculated structure factor
  - probability is complex Gaussian
- Only part of model structure factor is correct
  - plus random error

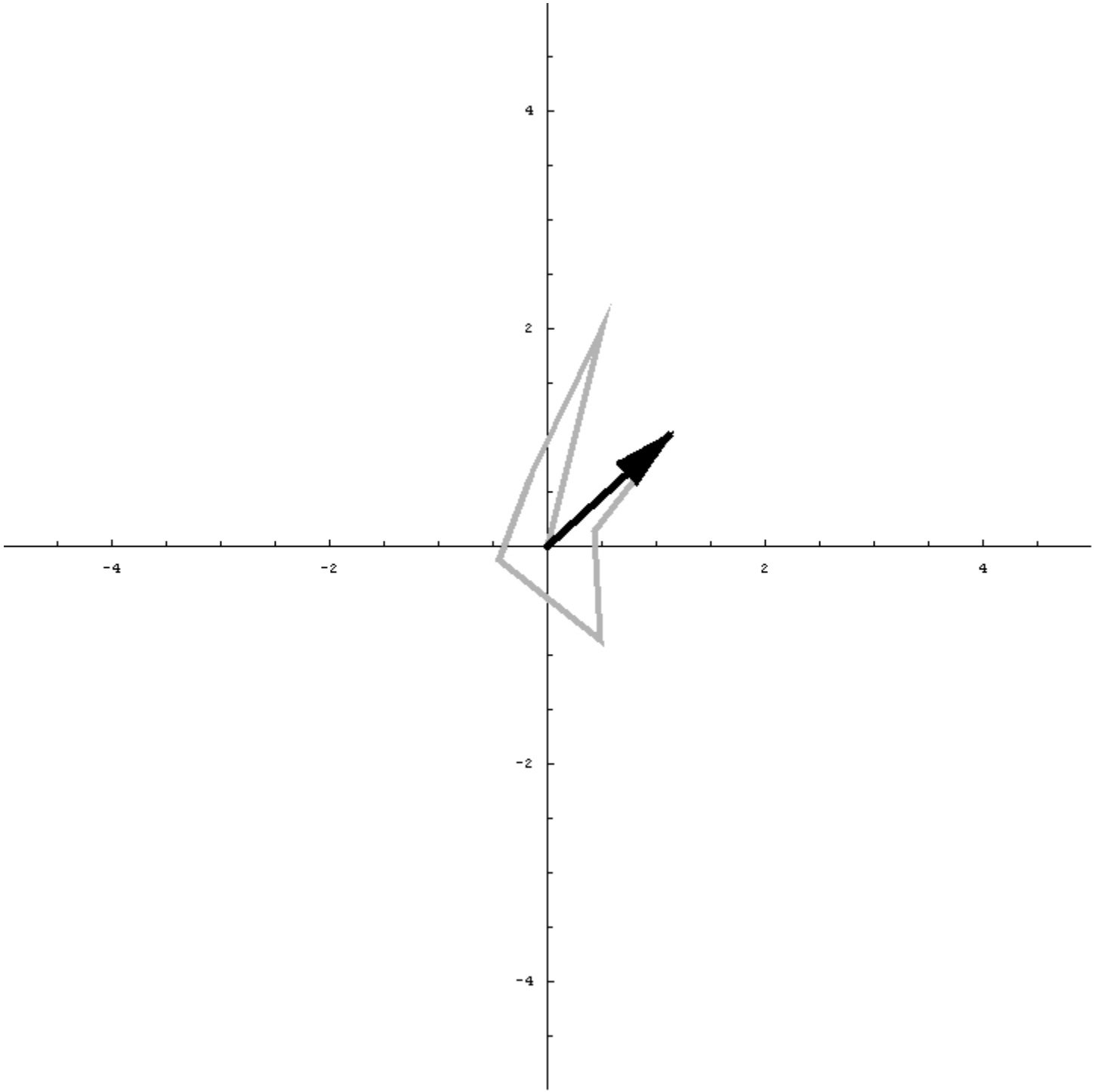


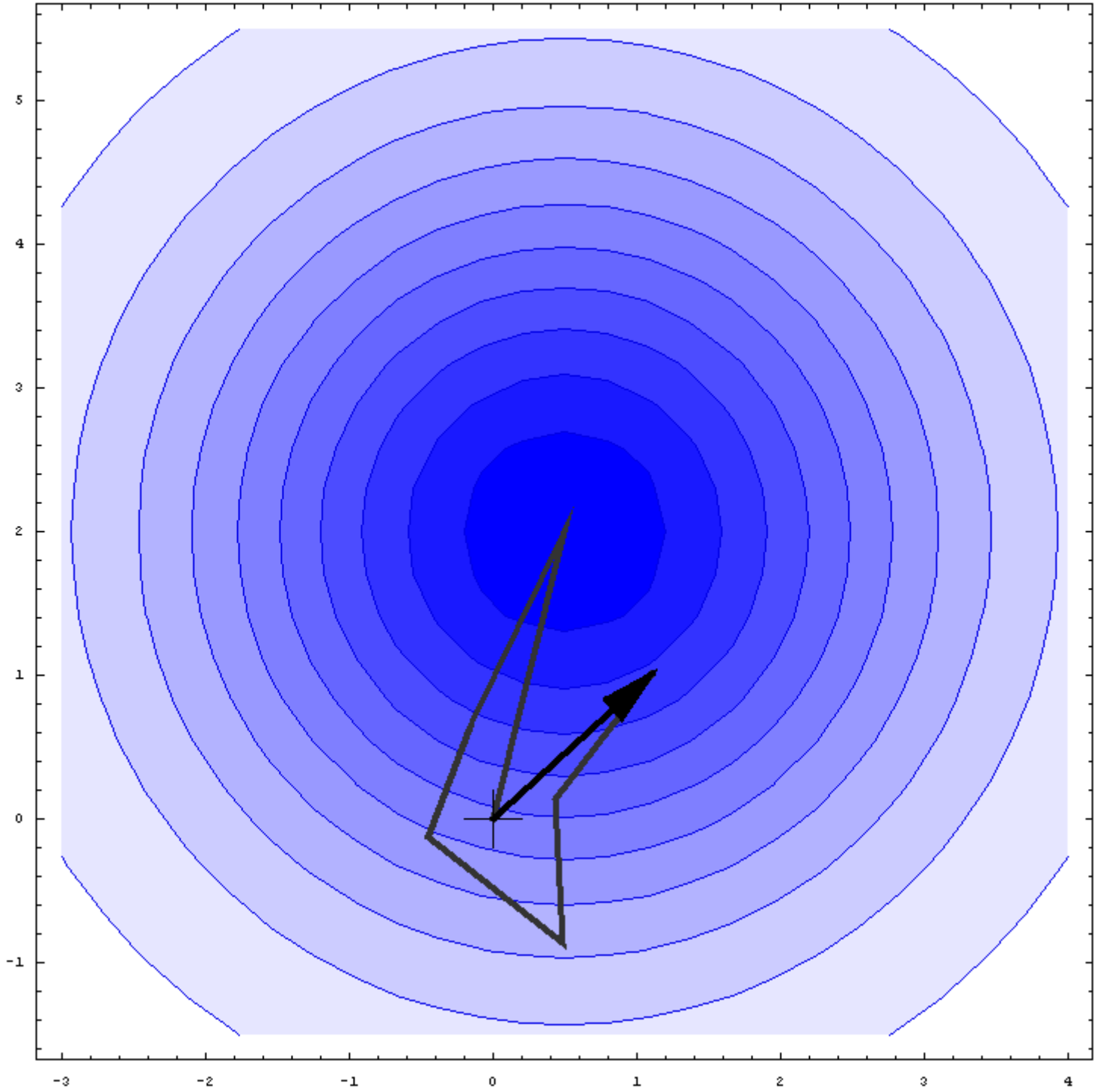
Rotation likelihood function

- What structure factors could be obtained from an oriented model?
  - add up contributions from molecules in unit cell, but unknown relative phase



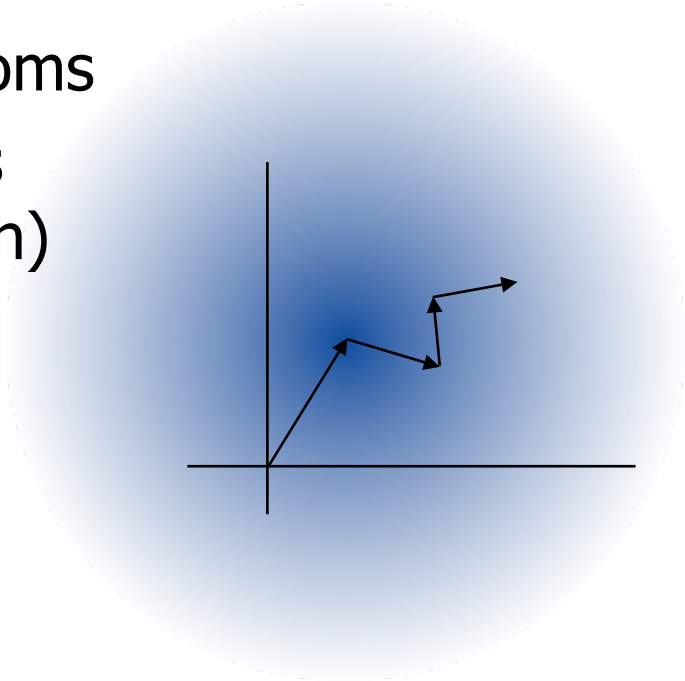



4
2
-4
-2
2
4
-2
-4


5
4
3
2
1
0
-1
-3
-2
-1
0
1
2
3
4

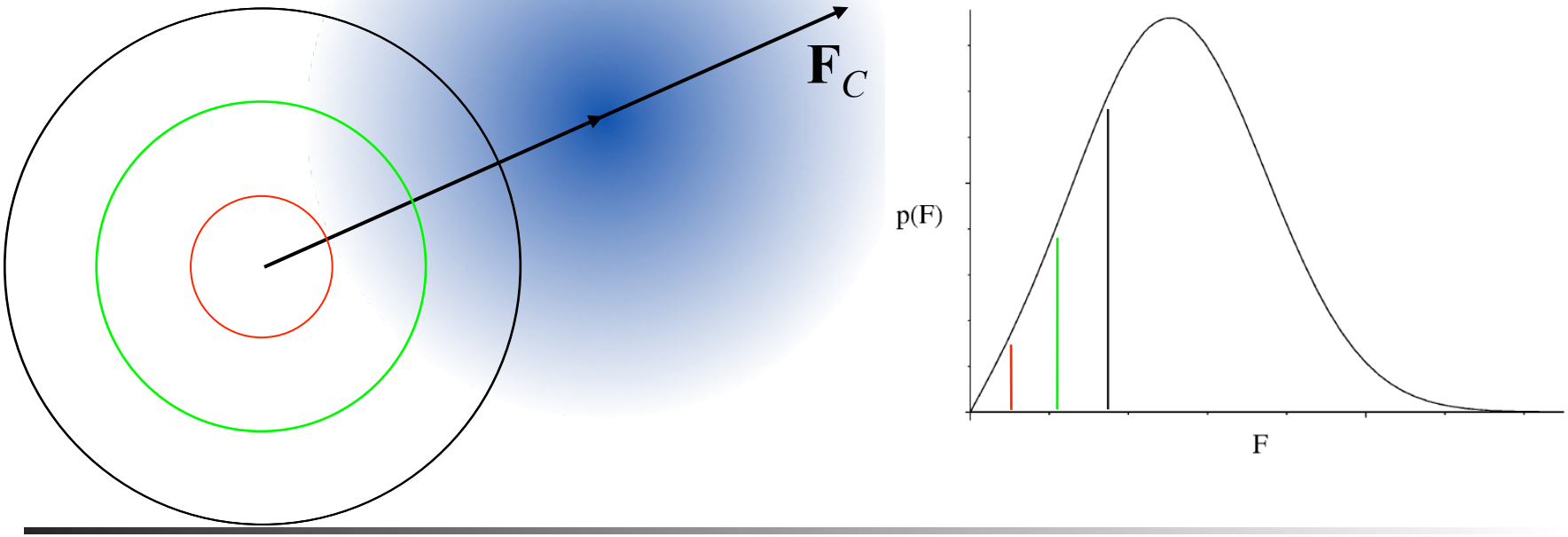
Molecular replacement likelihood function in *Phaser*

- Take biggest single contribution as $\mathbf{F}_C$
  - multiply by D
- Gaussian noise includes:
  - effects of model error, missing atoms
  - other symmetry-related molecules in cell (but not if translation search)



Amplitude probability distribution

- Have p($\mathbf{F}$), but data are |$\mathbf{F}$| so need p(|$\mathbf{F}$|)
- Integrate over unknown phase angle to get Rice (Luzzati, Sim, Srinivasan) distribution



Fast rotation and translation functions

- Full likelihood functions are expensive to evaluate
- Search orientations with likelihood-based fast rotation function
  - rescore plausible solutions with full rotation likelihood
- Search translations with likelihood-based fast translation function
  - rescore with full likelihood target
  - refine against full likelihood target

Correcting for anisotropy

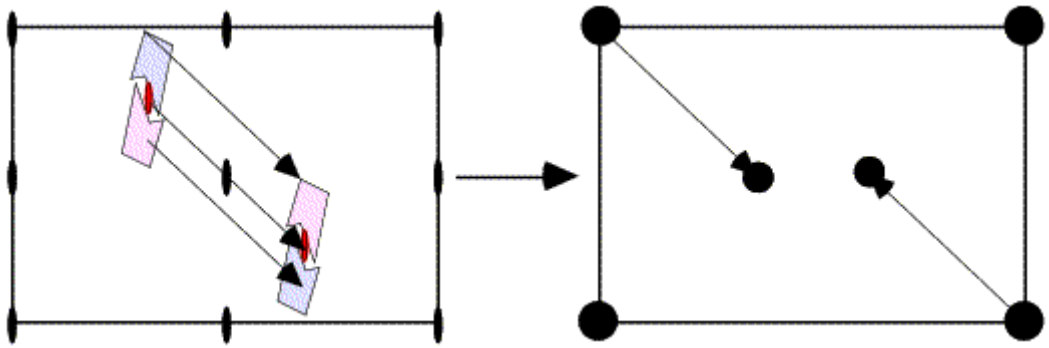
- Likelihood targets assume data are isotropic
- Correct anisotropy with anisotropic normalisation

$$\Sigma_N^{\mathrm{ANISO}} = \Sigma_N^{\mathrm{ISO}} \exp\left(-\left(\begin{array}{l}\beta_{HH}h^2 + \beta_{KK}k^2 + \beta_{LL}l^2 \\ +\beta_{HK}hk + \beta_{HL}hl + \beta_{KL}kl\end{array}\right)\right)$$

- Refine β parameters so that normalised data match Wilson distribution

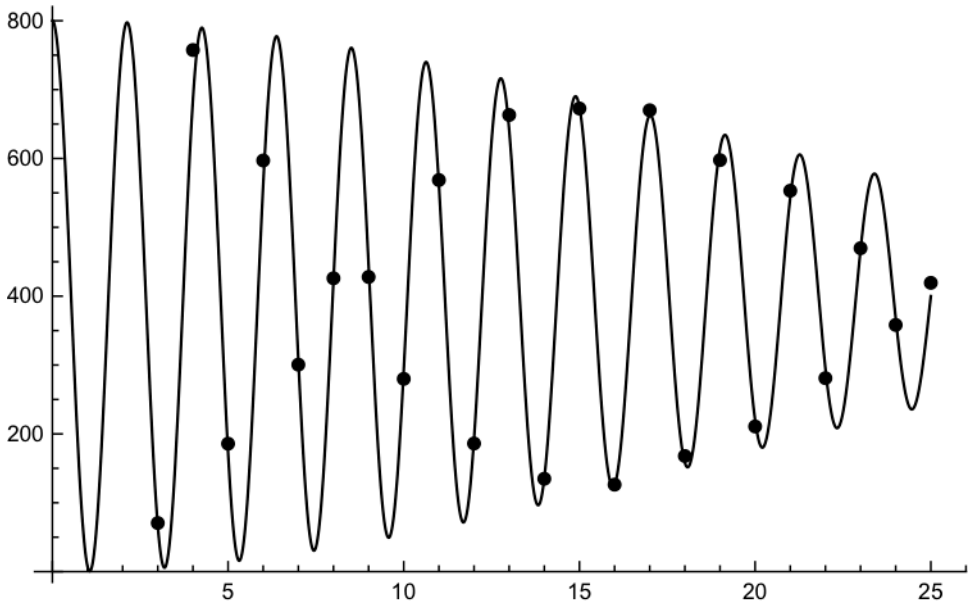
Translational NCS

- Found in about 8% of PDB entries
- NCS often parallel to crystallographic symmetry
  - combination gives translational NCS (tNCS)



Accounting for translational NCS

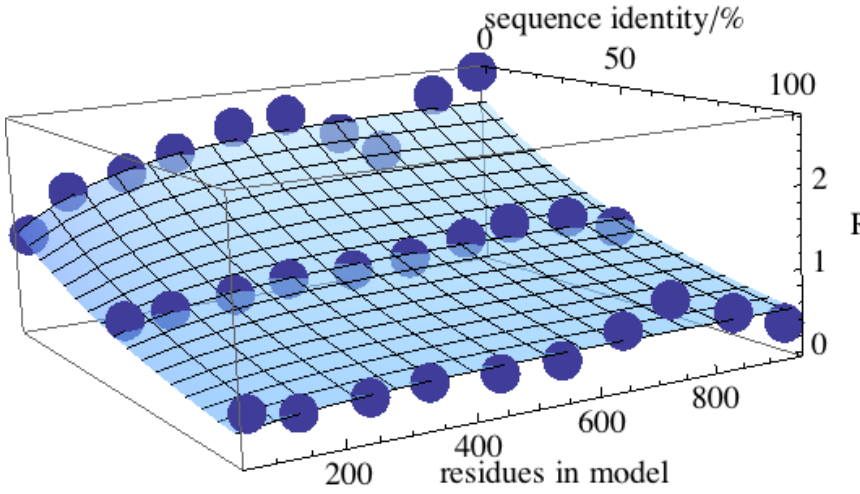
- Model effect of translation combined with small rotation and random differences between copies

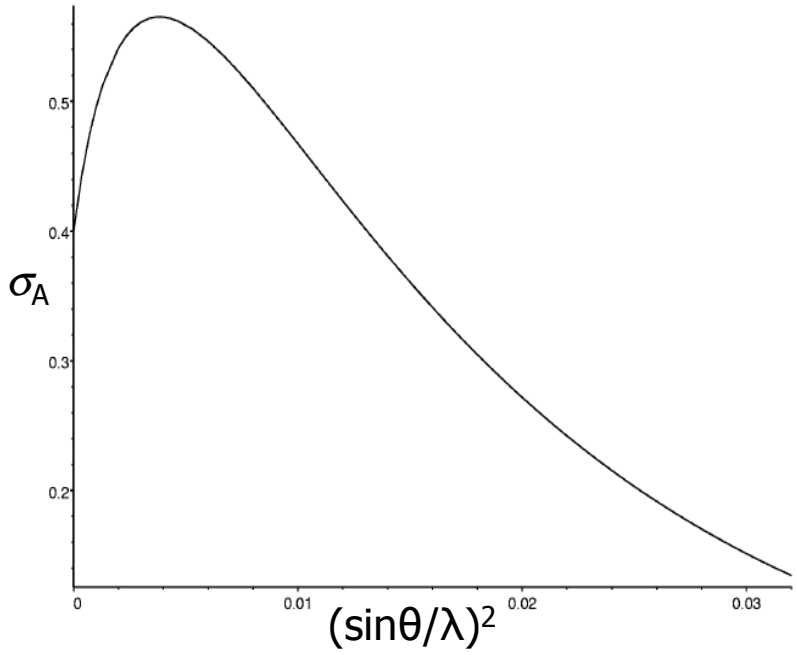


Hyp-1:
Sliwiak, Jaskolski, Dauter, McCoy, Read
(unpublished)

A priori estimate of model quality

- Model quality translates into parameter σ_A
  - depends on completeness, disordered solvent, errors
- Optimal RMS error from database of MR trials

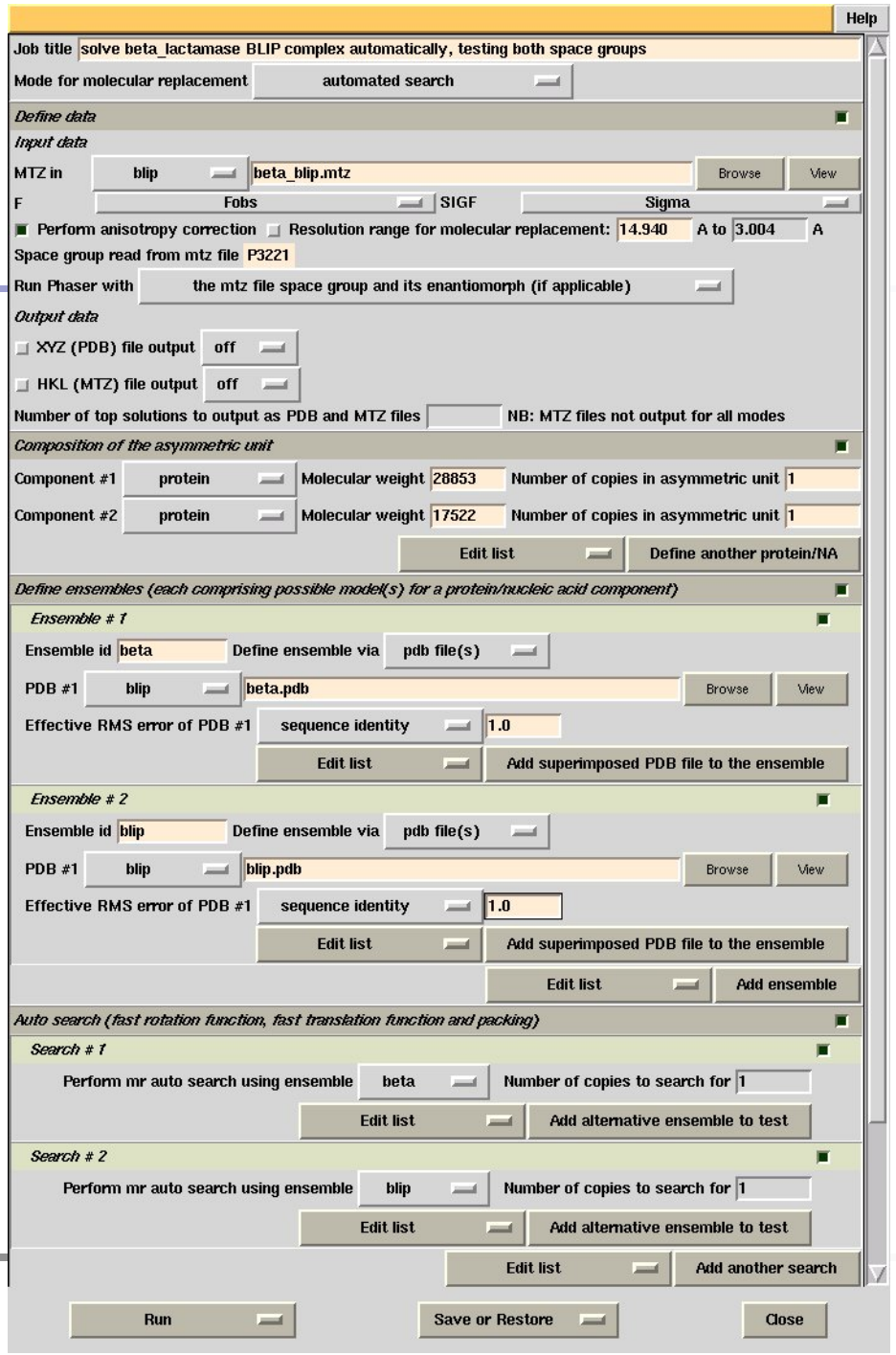


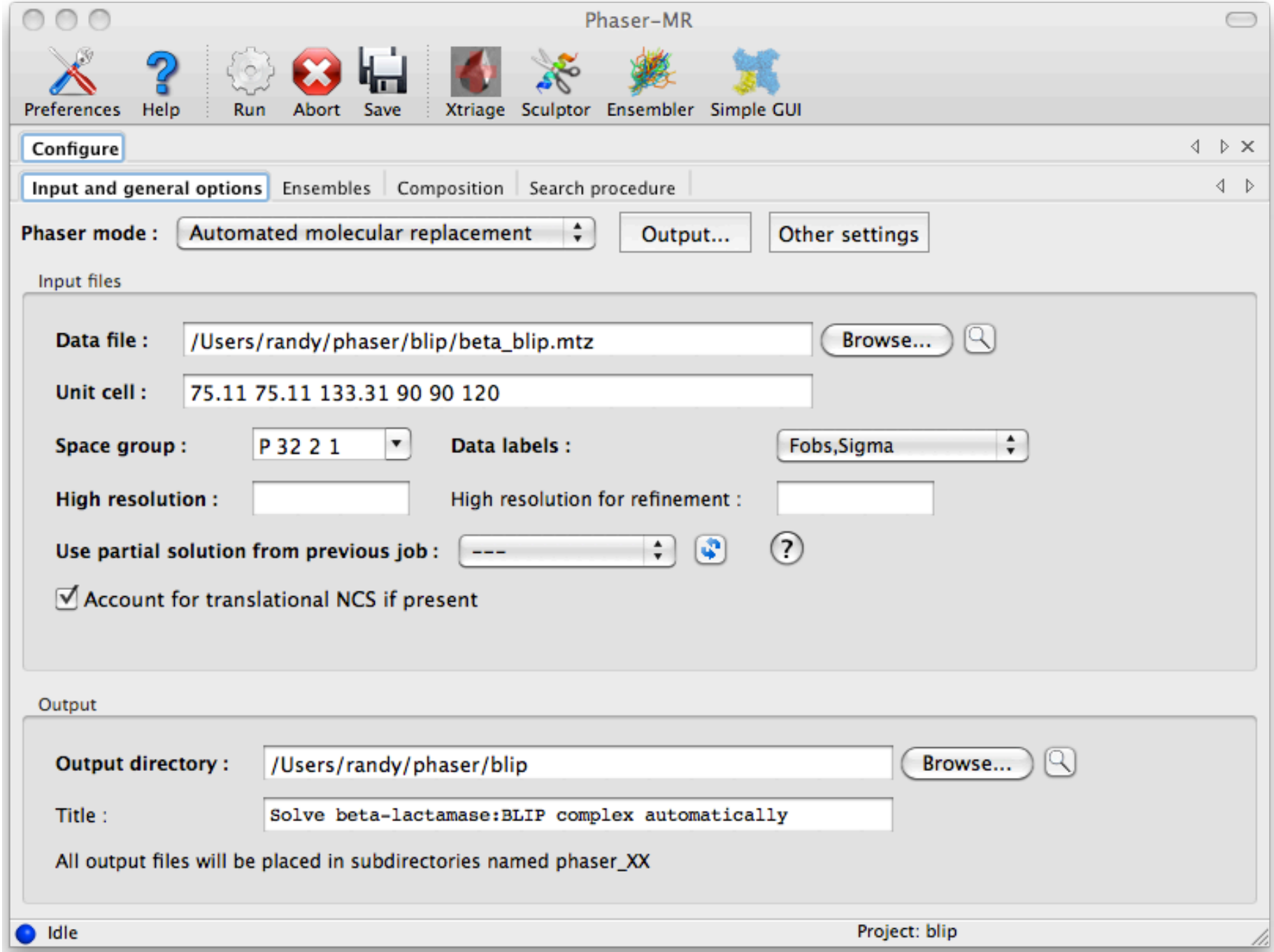


Ways to run *Phaser*

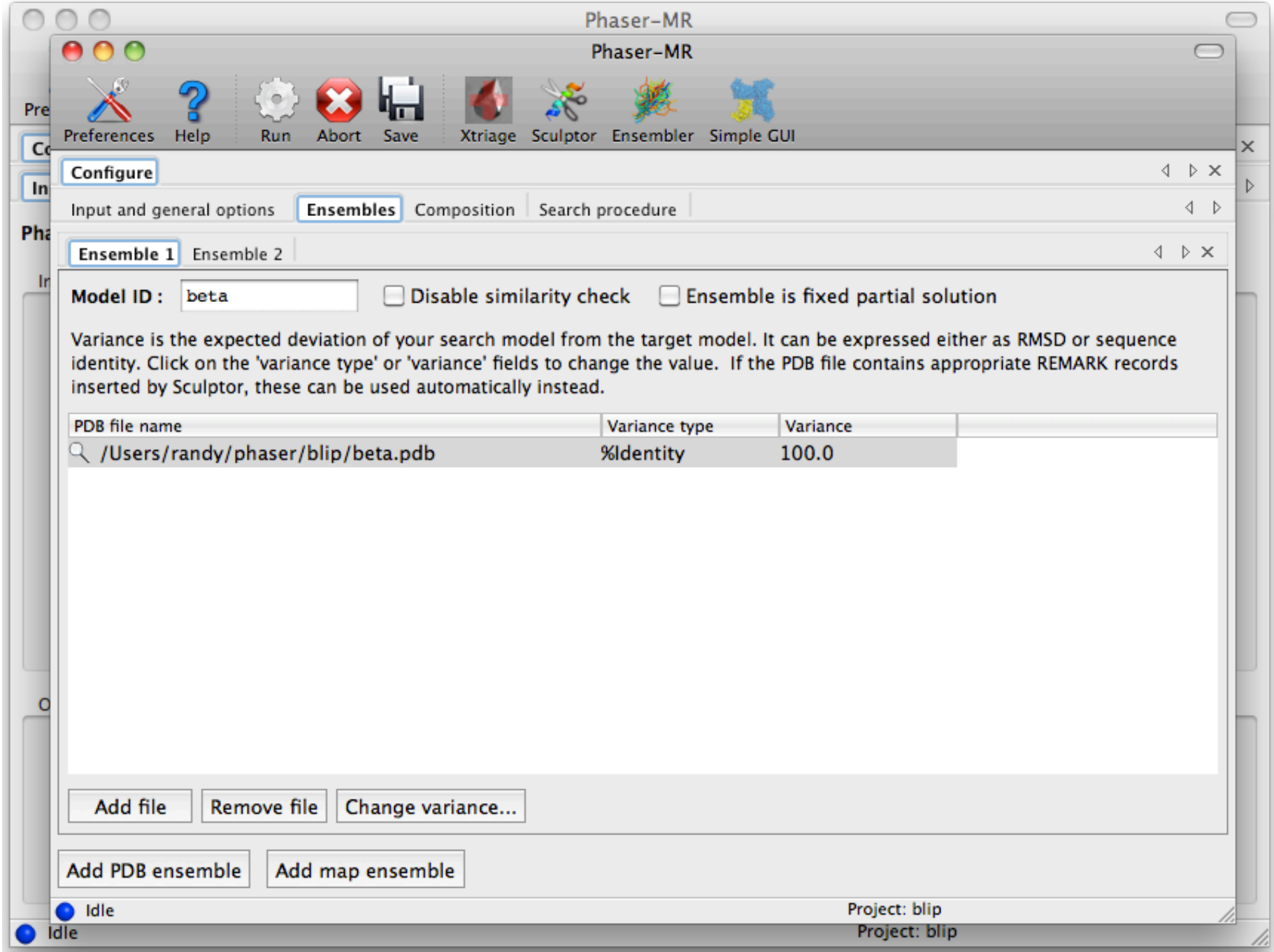
- Distributed with Phenix and CCP4
- GUIs available in both packages
- Run from keyword scripts or Python scripts

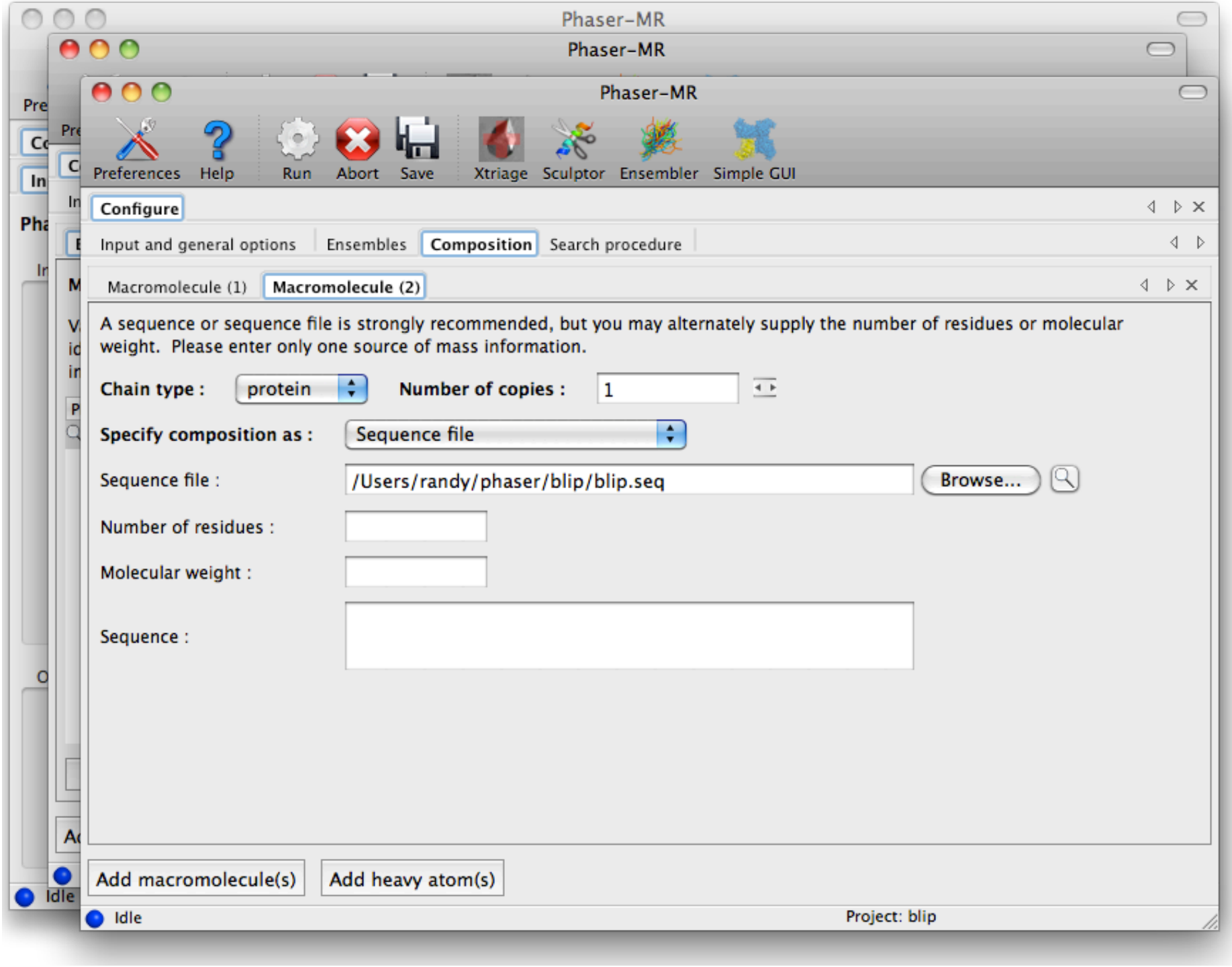
ccp4i GUI for *Phaser*

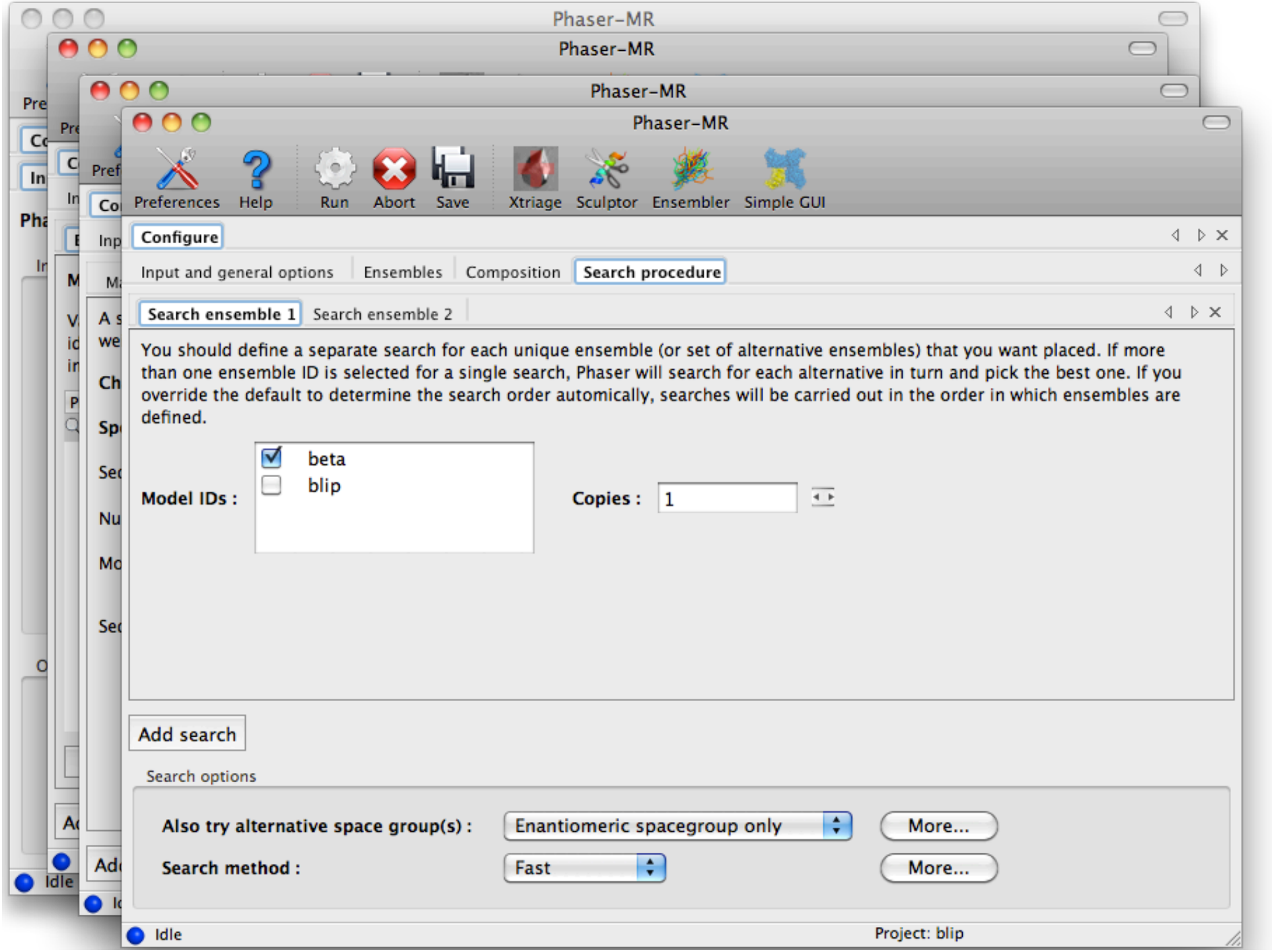



Phaser-MR
Preferences
Help
Run
Abort
Save
Xtriage
Sculptor
Ensembler
Simple GUI
Configure
Input and general options
Ensembles
Composition
Search procedure
Phaser mode :
Automated molecular replacement
Output...
Other settings
Input files
Data file :
/Users/randy/phaser/blip/beta_blip.mtz
Browse...
Unit cell :
75.11 75.11 133.31 90 90 120
Space group :
P 32 2 1
Data labels :
Fobs,Sigma
High resolution :
High resolution for refinement :
Use partial solution from previous job :

Account for translational NCS if present
Output
Output directory :
/Users/randy/phaser/blip
Browse...
Title :
Solve beta-lactamase:BLIP complex automatically
All output files will be placed in subdirectories named phaser_XX
Idle
Project: blip

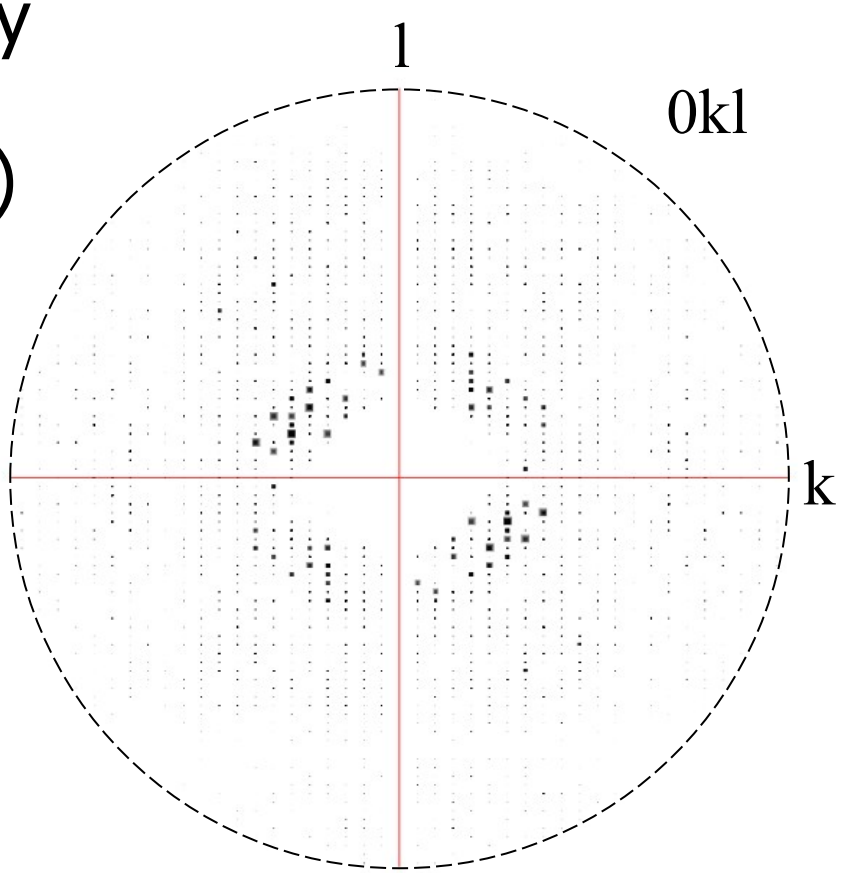

Phaser-MR
Phaser-MR
Preferences
Help
Run
Abort
Save
Xtriage
Sculptor
Ensembler
Simple GUI
Configure
Input and general options
Ensembles
Composition
Search procedure
Ensemble 1
Ensemble 2
Model ID :
beta
Disable similarity check
Ensemble is fixed partial solution
Variance is the expected deviation of your search model from the target model. It can be expressed either as RMSD or sequence identity. Click on the 'variance type' or 'variance' fields to change the value. If the PDB file contains appropriate REMARK records inserted by Sculptor, these can be used automatically instead.
PDB file name
Variance type
Variance
/Users/randy/phaser/blip/beta.pdb
%Identity
100.0
Add file
Remove file
Change variance...
Add PDB ensemble
Add map ensemble
Idle
Project: blip
Idle
Project: blip


Phaser-MR
Phaser-MR
Phaser-MR
Preferences
Help
Run
Abort
Save
Xtriage
Sculptor
Ensembler
Simple GUI
Configure
Input and general options
Ensembles
Composition
Search procedure
Macromolecule (1)
Macromolecule (2)
A sequence or sequence file is strongly recommended, but you may alternately supply the number of residues or molecular weight. Please enter only one source of mass information.
Chain type :
protein
Number of copies :
1
Specify composition as :
Sequence file
Sequence file :
/Users/randy/phaser/blip/blip.seq
Browse...
Number of residues :
Molecular weight :
Sequence :
Add macromolecule(s)
Add heavy atom(s)
Idle
Project: blip


Phaser-MR
Preferences
Help
Run
Abort
Save
Xtriage
Sculptor
Ensembler
Simple GUI
Configure
Input and general options
Ensembles
Composition
Search procedure
Search ensemble 1
Search ensemble 2
You should define a separate search for each unique ensemble (or set of alternative ensembles) that you want placed. If more than one ensemble ID is selected for a single search, Phaser will search for each alternative in turn and pick the best one. If you override the default to determine the search order automically, searches will be carried out in the order in which ensembles are defined.
Model IDs :
beta
blip
Copies :
1
Add search
Search options
Also try alternative space group(s) :
Enantiomeric spacegroup only
More...
Search method :
Fast
More...
Idle
Project: blip

β-lactamase:BLIP complex

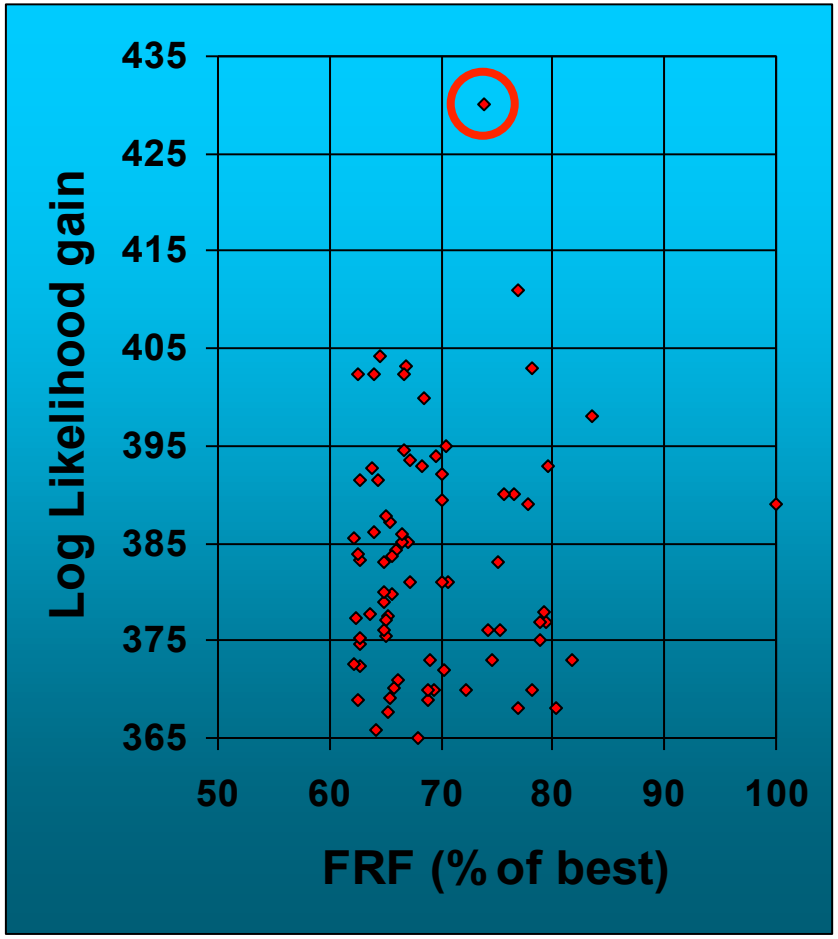
- Solved with great difficulty using AMoRe (Strynadka, James, Alzari)
- β-lactamase
  - 62% of the structure
  - easy to find
- BLIP
  - 38% of the structure
  - hard to find
- Anisotropic diffraction



β-lactamase:BLIP complex before *Phaser*

- Crowther target
- β-lactamase not used
- Anisotropic data

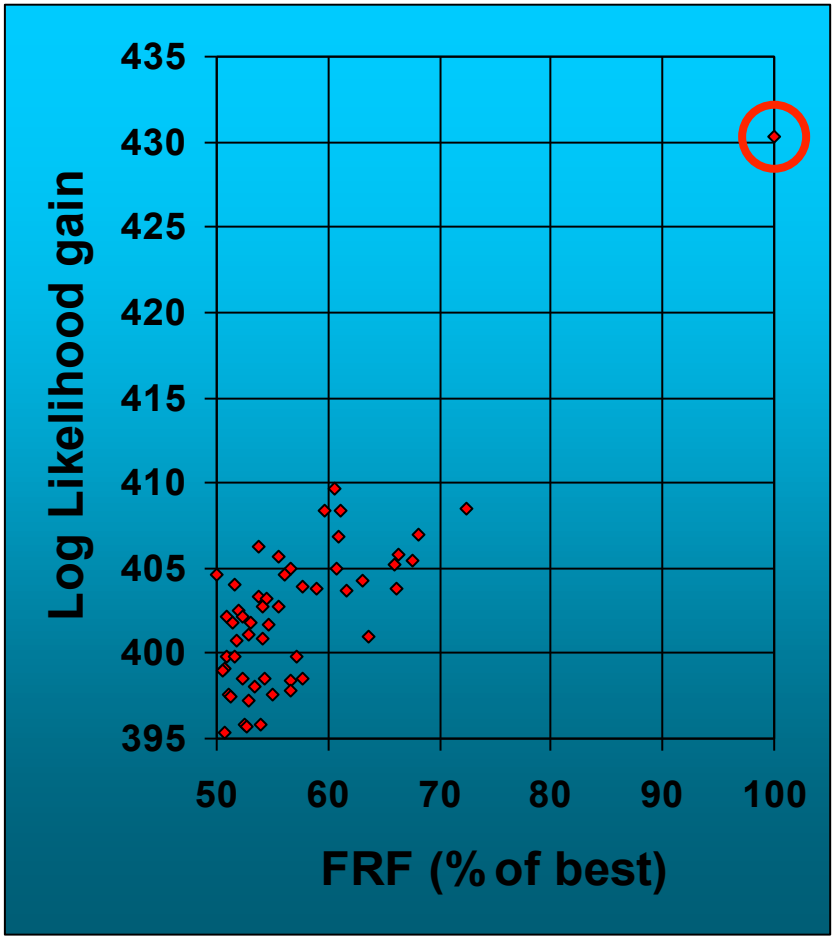
- Correct peak in noise



β-lactamase:BLIP complex after *Phaser*

- LERF1 target
- fix β-lactamase
- Anisotropy corrected

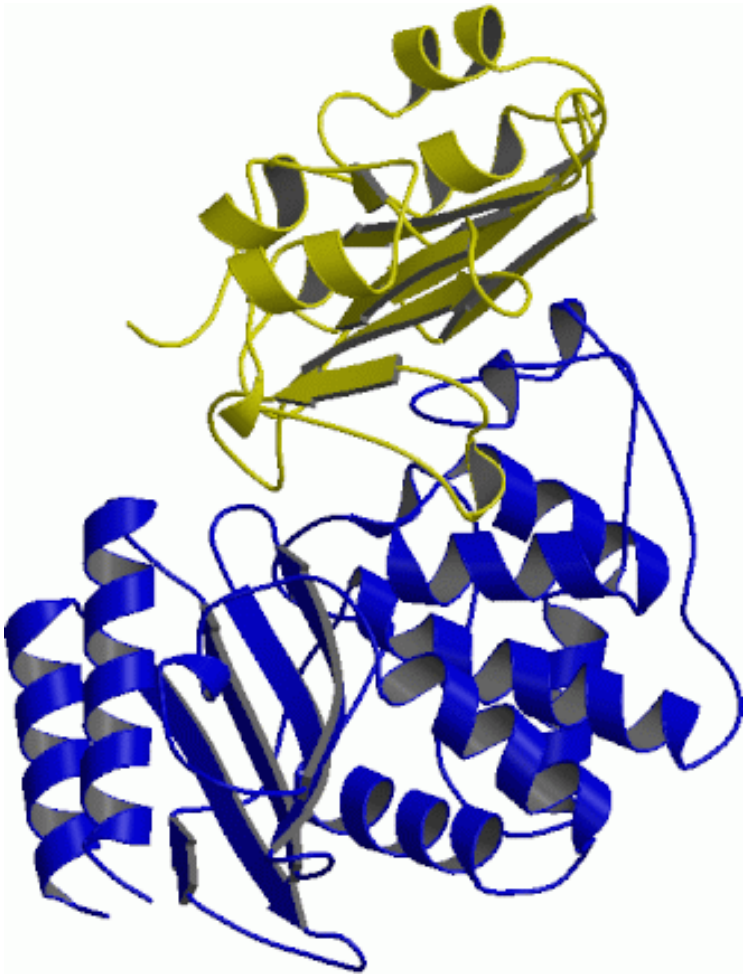
- Clear peak
- Result in minutes



β-lactamase:BLIP complex after *Phaser*

- LERF1 target
- fix β-lactamase
- Anisotropy corrected

- Clear peak
- Result in minutes

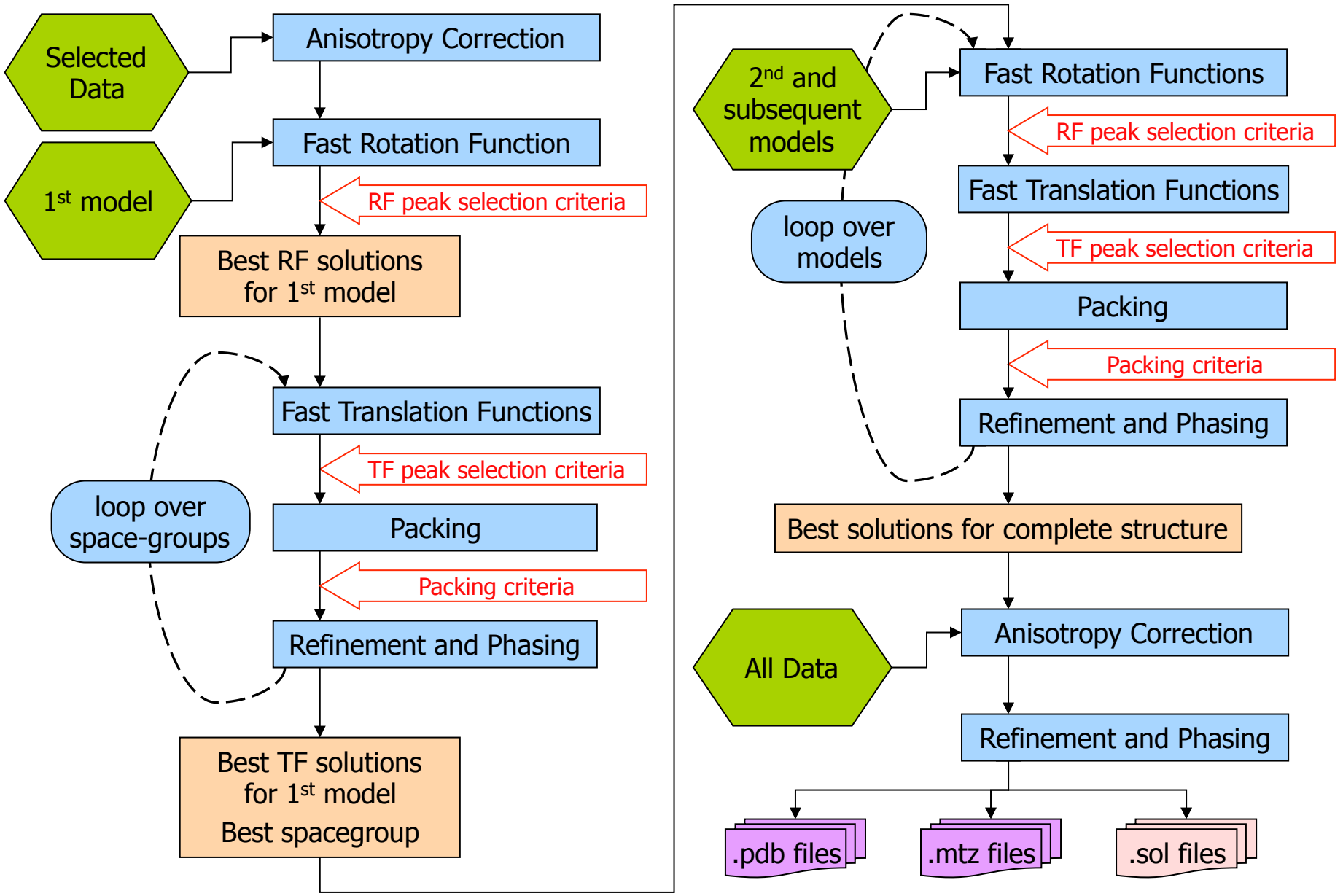


Likelihood and automation

- Automated decisions require reliable scores
- Likelihood provides absolute score
  - compare different models
  - compare different space groups
  - likelihood should increase for better model
    - more accurate, more complete or more detailed
- Other aspects of automation
  - check packing
  - keep track of potential partial solutions
  - add multiple components

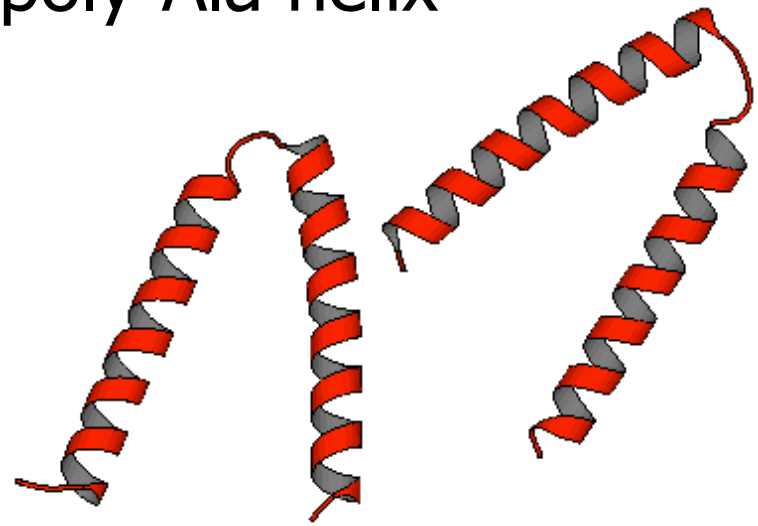
Automation in *Phaser*

- MR_AUTO mode
  - searches over possible space-groups
  - checks potential solutions for packing
  - refines solutions away from search grid to optimal orientation and position
  - uses parts of the structure already found to bootstrap the entire solution
  - amalgamates compatible partial solutions
- Protocol fine-tuned with difficult MR problems
- MRage: full pipeline (available in Phenix)

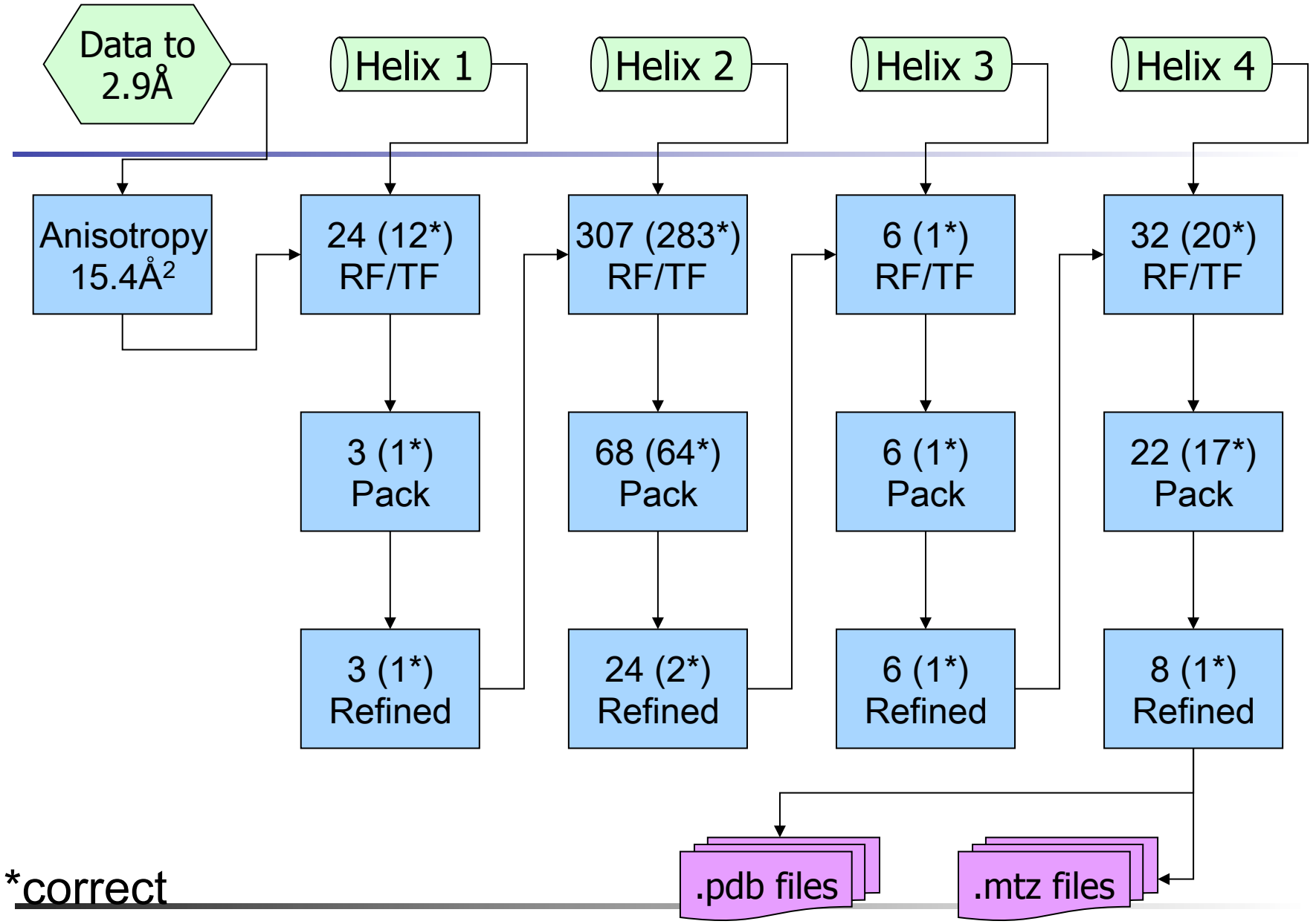

Selected Data
1st model
Anisotropy Correction
Fast Rotation Function
RF peak selection criteria
Best RF solutions for 1st model
Fast Translation Functions
TF peak selection criteria
Packing
Packing criteria
Refinement and Phasing
loop over space-groups
Best TF solutions for 1st model
Best spacegroup
2nd and subsequent models
loop over models
Fast Rotation Functions
RF peak selection criteria
Fast Translation Functions
TF peak selection criteria
Packing
Packing criteria
Refinement and Phasing
Best solutions for complete structure
All Data
Anisotropy Correction
Refinement and Phasing
.pdb files
.mtz files
.sol files

A31P mutant of ROP: four helix bundle

- Originally solved by 23-dimensional Monte Carlo search with four copies of poly-Ala helix
  - space group C2
  - helix = 15% of protein
  - Glykos & Kokkinidis (2003)



- Can be solved in minutes by *Phaser*


Data to 2.9Å
Helix 1
Helix 2
Helix 3
Helix 4
Anisotropy 15.4Å2
24 (12*) RF/TF
307 (283*) RF/TF
6 (1*) RF/TF
32 (20*) RF/TF
3 (1*) Pack
68 (64*) Pack
6 (1*) Pack
22 (17*) Pack
3 (1*) Refined
24 (2*) Refined
6 (1*) Refined
8 (1*) Refined
.pdb files
.mtz files
*correct

Arcimboldo

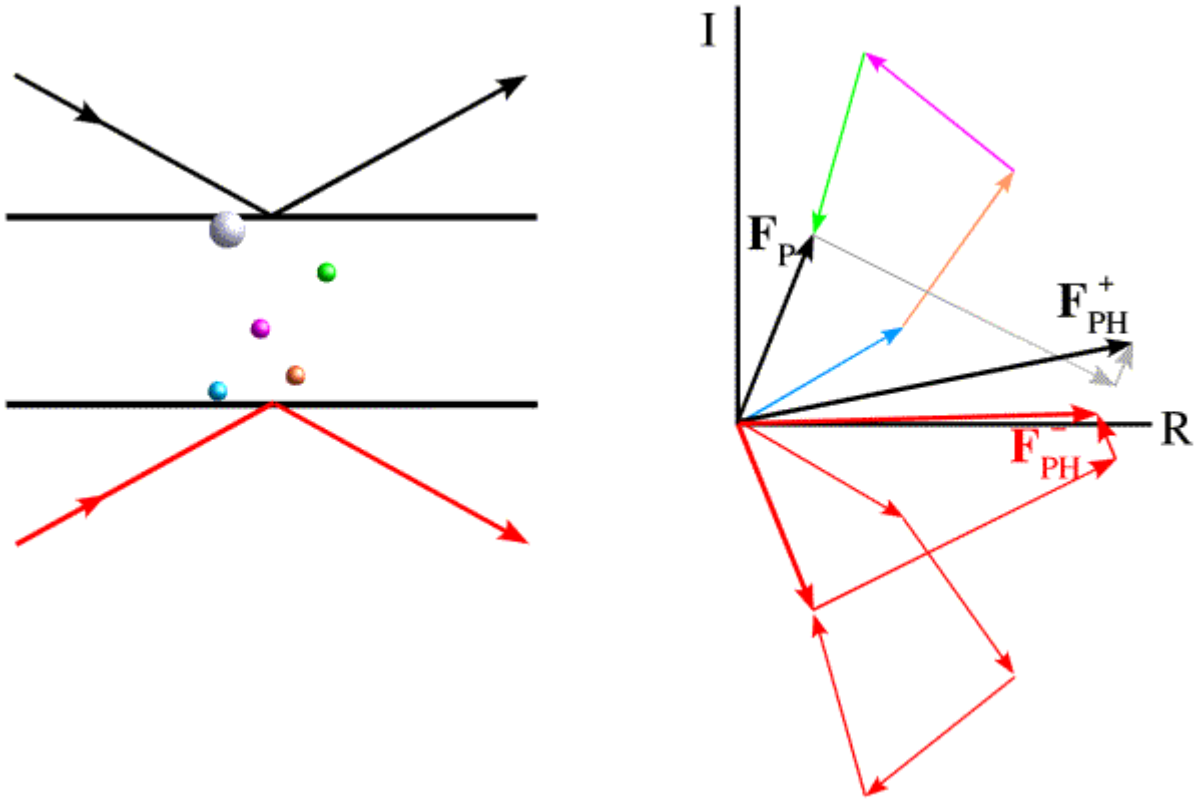
- Rodríguez, Grosse, Himmel, González, de Ilarduya, Becker, Sheldrick & Usón, "Crystallographic *ab initio* protein structure solution below atomic resolution", *Nature Methods* **6**: 651-653 (2009)
- *Phaser* and *SHELXE*

Practical aspects of MR in *Phaser*

- Provide information about model quality
  - estimated RMS error to calibrate σ_A curve
- Provide information about cell content
  - sequence, molecular weight, percent solvent…
  - used to determine model completeness
- Consider possibility of conformational change
  - alternative models
  - search with isolated domains

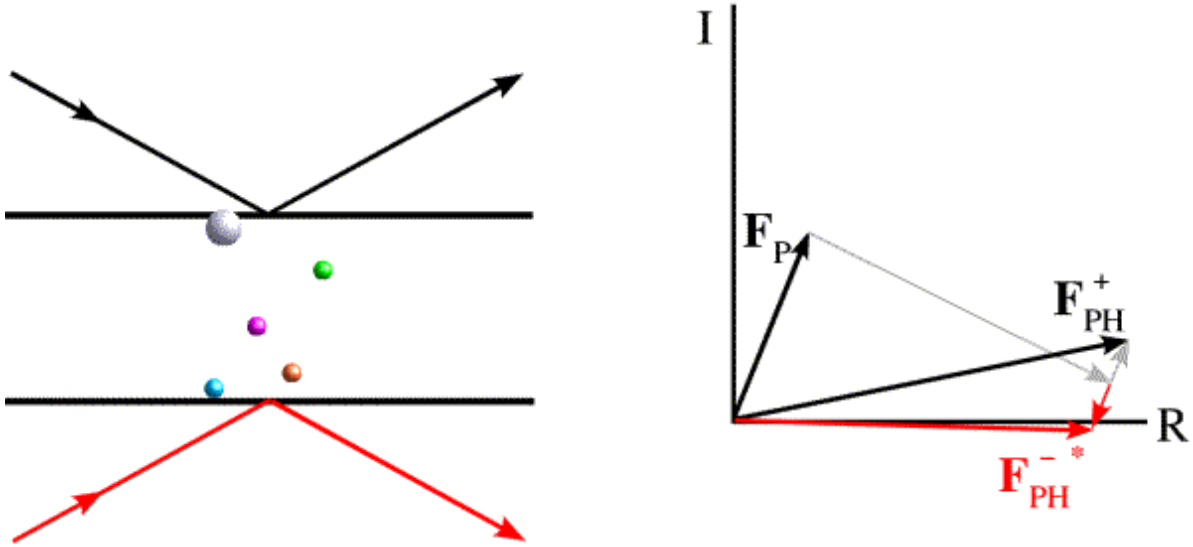
Combining MR and SAD information

- SAD: single-wavelength anomalous diffraction

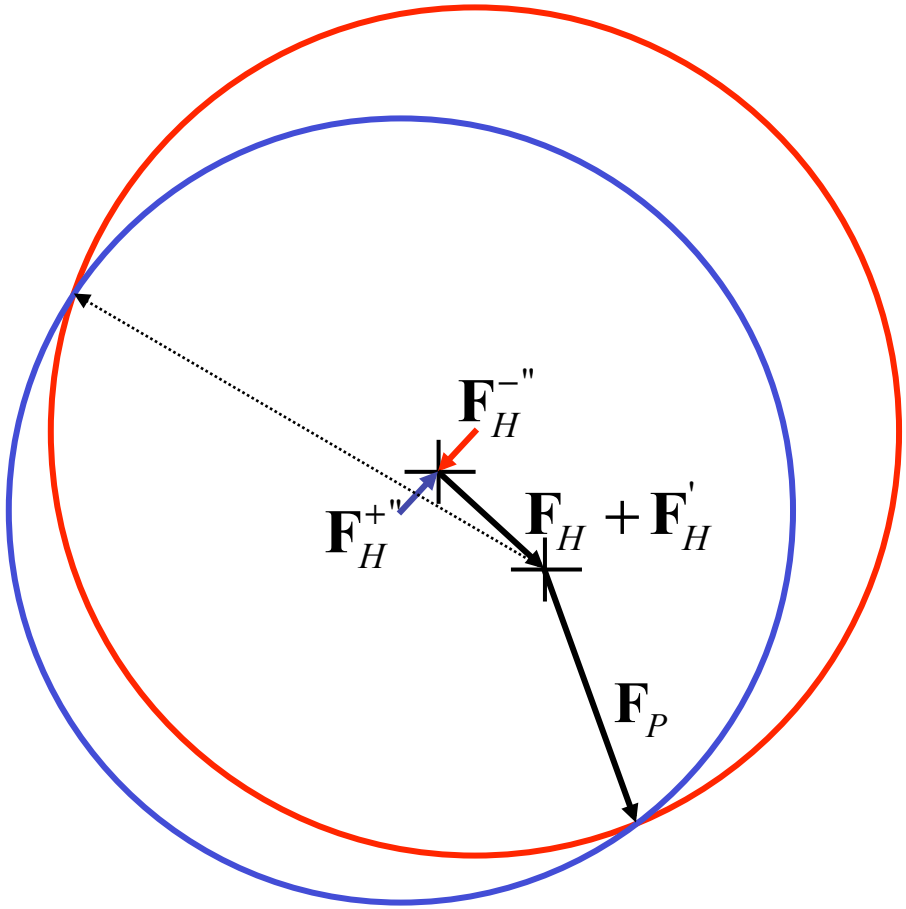


Combining MR and SAD information

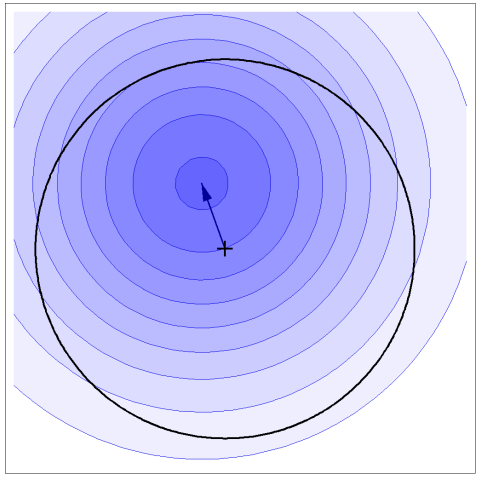
- SAD: single-wavelength anomalous diffraction

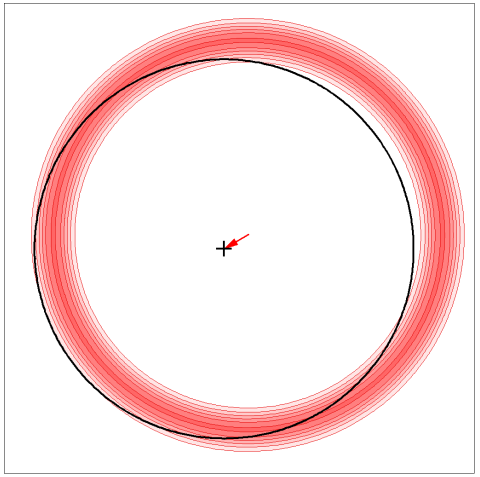


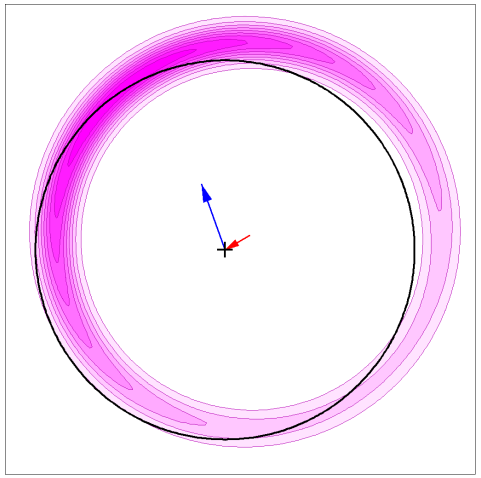
Harker construction for SAD phasing

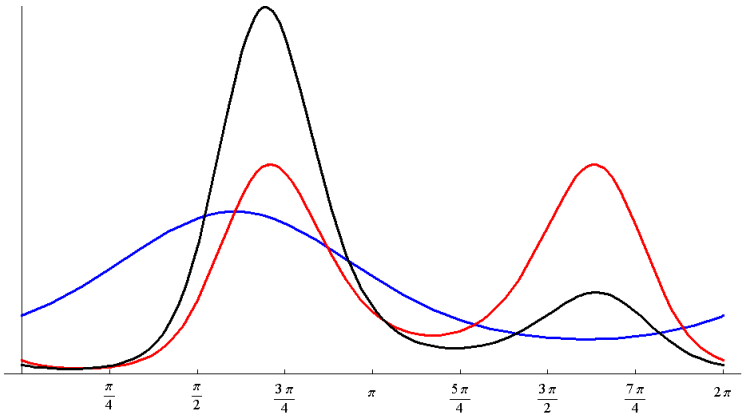


SAD: likelihood based on joint probabilities









SAD log-likelihood gradient (LLG) map

- Compute derivative of log-likelihood with respect to heavy atom structure factor
- Fourier transform gives map of where likelihood target would like to see changes in anomalous scatterer model
- Very sensitive to minor sites
  - picks up sites identified as water molecules in refined structures determined by halide soaks

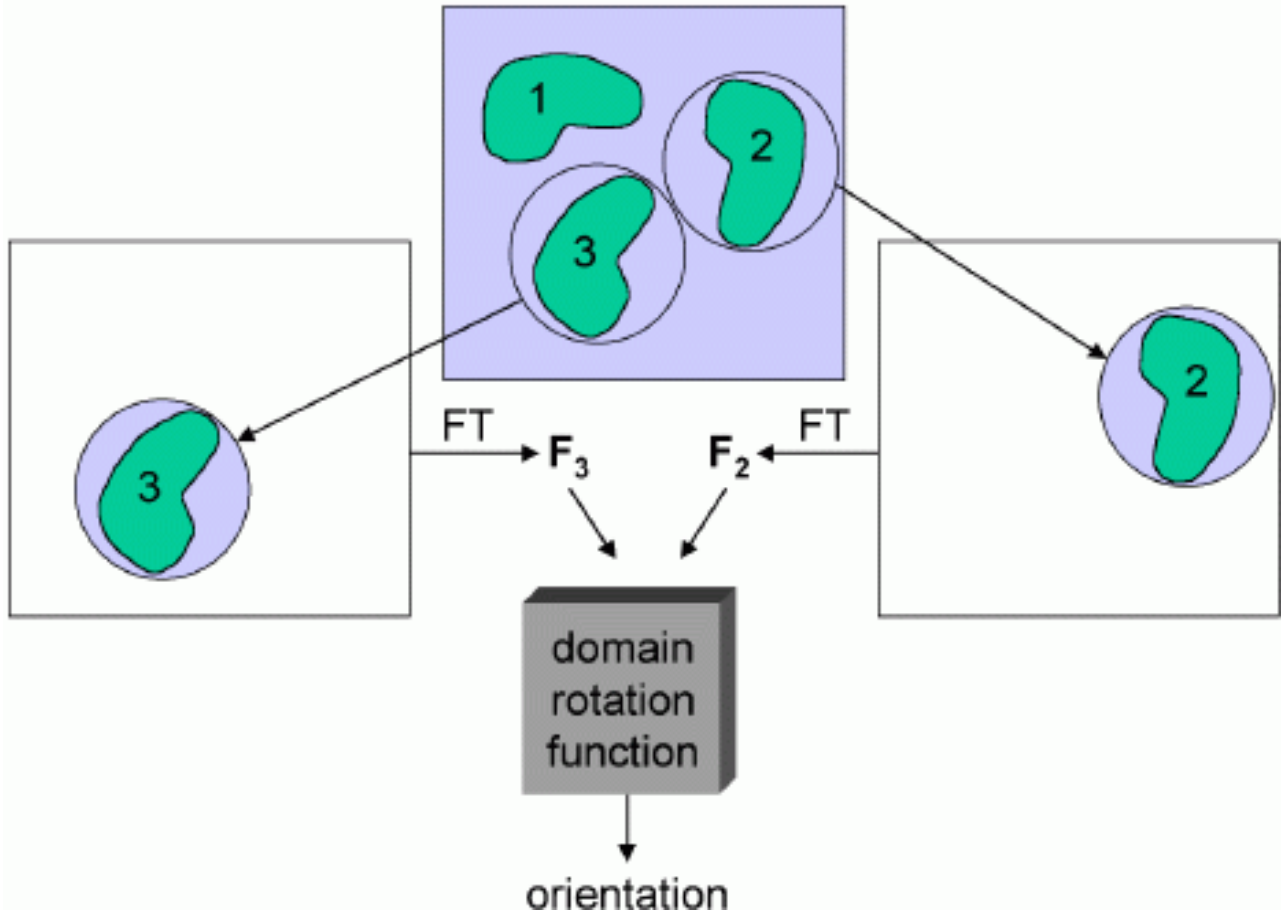
Combining MR and SAD

- Solve structure by molecular replacement
- Use SAD likelihood to add anomalous scatterers
- Phase information automatically combined

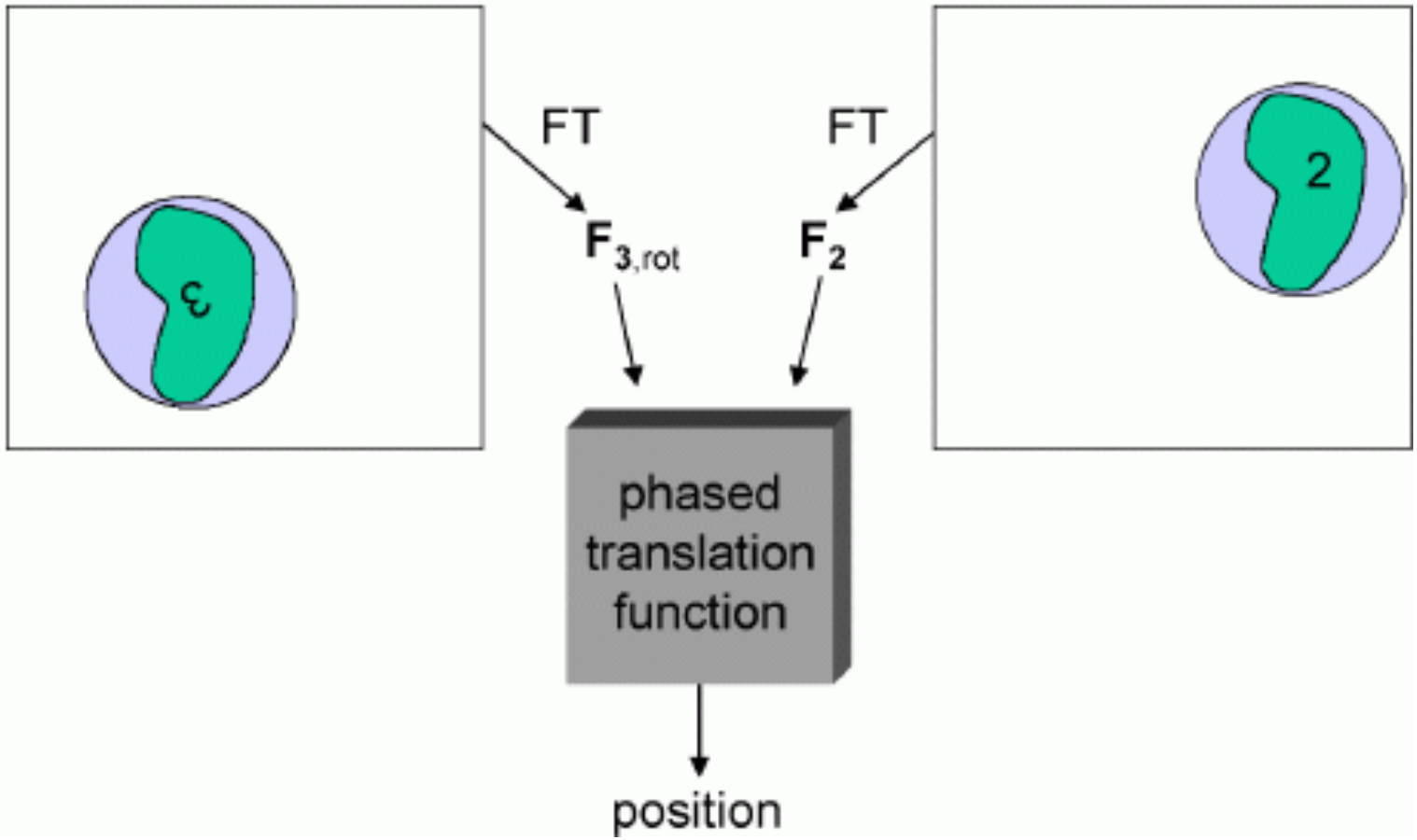
Real-space molecular replacement

- Use phase information in two ways:
  - use electron density as model
    - calculate structure factors from isolated density, then proceed as with atomic model
    - possible in *Phaser*
  - fit model into electron density
    - "domain rotation function"
    - "phased translation function"
    - not yet possible in *Phaser*

Domain rotation function



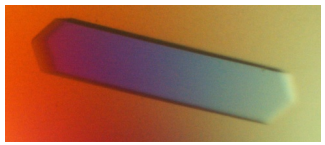
Phased translation function



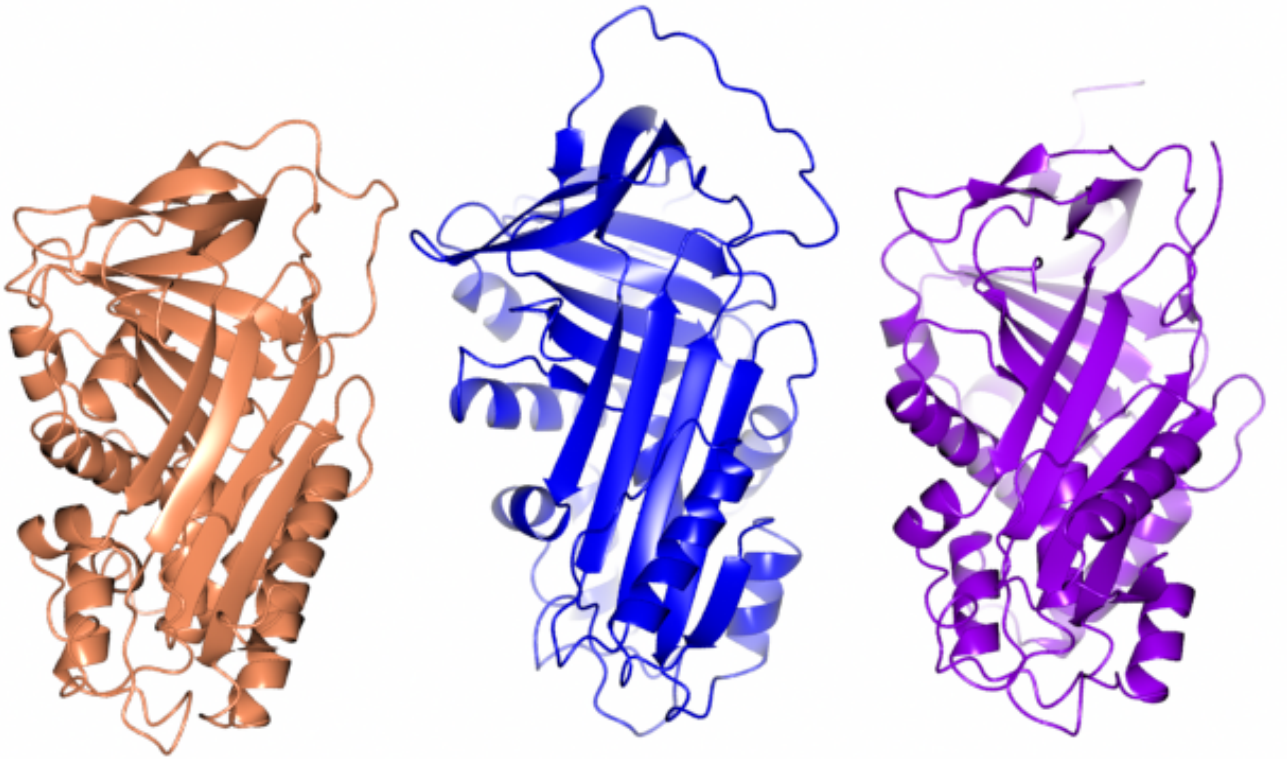
Angiotensinogen crystals

- Human: 1 crystal form
  - 3.3Å, 1 copy
- Rat: 2 crystal forms
  - 2.8Å, 2 copies
  - 3.15Å, 2 copies
- Mouse: 2 crystal forms
  - 2.1Å, 1 copy
  - 2.95Å, 4 copies

Human angiotensinogen: molecular replacement



human

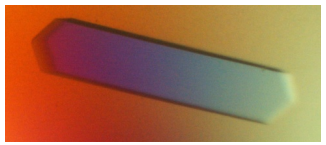


heparin cofactor II
(20% identical)

α_1-antitrypsin
(21% identical)

thyroxine-binding globulin
(20% identical)

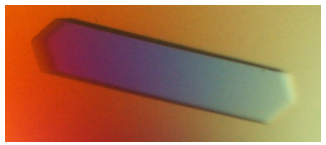
Human angiotensinogen: molecular replacement



human

Ensemble

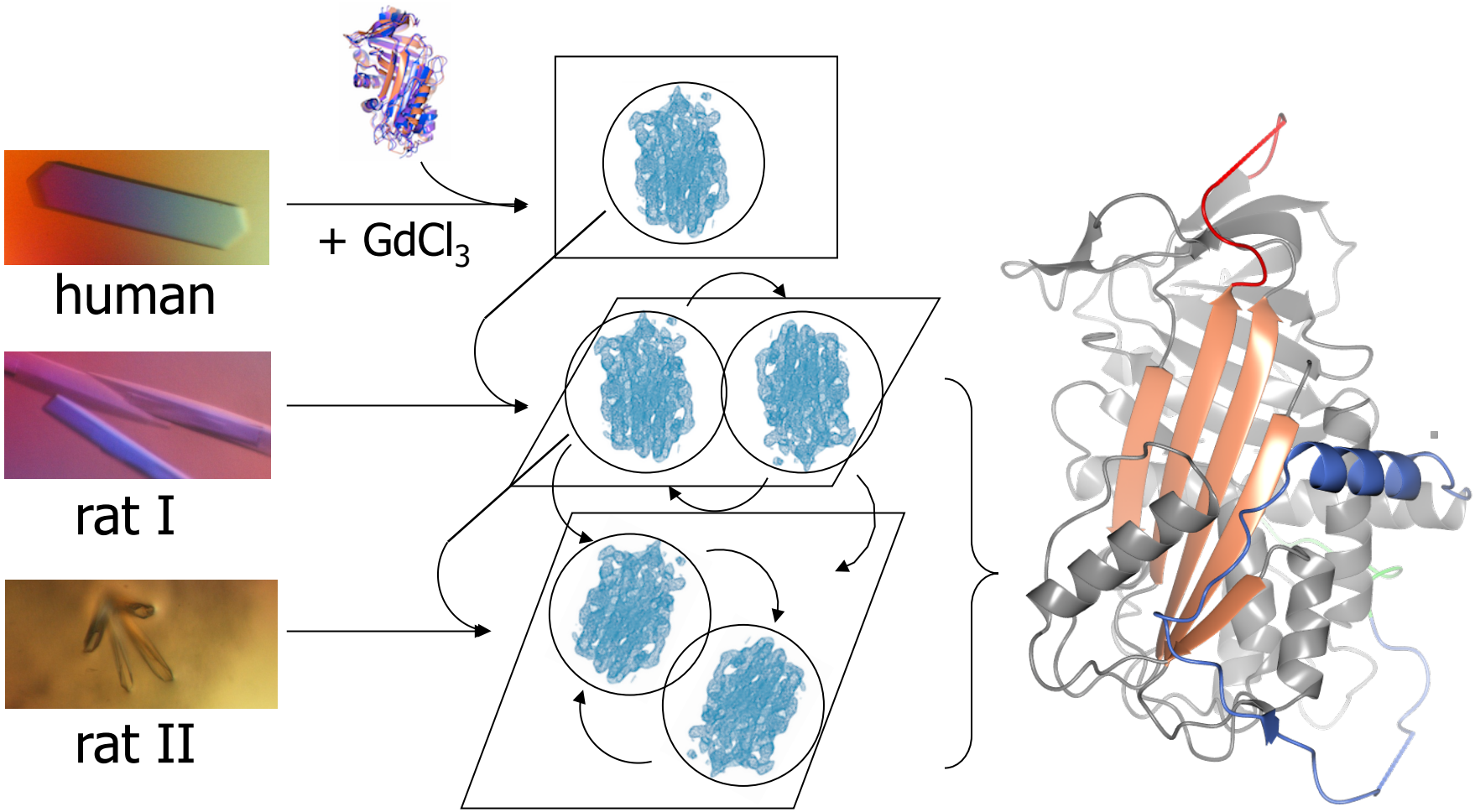
Human angiotensinogen: molecular replacement



human

Trimmed ensemble

Solving angiotensinogen structures



Solving *Drosophila* GST2 (1M0U)

- Difficult structure from Bogos Agianian (Piet Gros)
- Find one of two copies with ensemble of 3 structures (28-30% identity)
  - search for second copy fails
- Find second copy as density from first
  - this succeeds

Phaser Tutorials

- http://www-structmed.cimr.cam.ac.uk/phaser/tutorial
- MR tutorials
  - TOXD with separate models or ensemble
  - complex of β-lactamase with BLIP
- MR+SAD tutorial
  - Solve hen egg-white lysozyme with goat α-lactalbumin, then use LLG maps to find intrinsic anomalous scatterers (S and Cl), thereby improving phase information
    - anomalous differences are too poor to find substructure without MR model

Acknowledgments

- Development of *Phaser*
  - Airlie McCoy, Laurent Storoni, Gabor Bunkoczi, Robert Oeffner
- PHENIX collaboration
  - Ralf Grosse-Kunstleve, Nigel Moriarty, Paul Adams
  - Tom Terwilliger

Funding


wellcome trust


NATIONAL INSTITUTES OF HEALTH


British Heart Foundation


CCP4